



Restrictiveness and guidance in support systems

Paul Goodwin^{a,*}, Robert Fildes^b, Michael Lawrence^c, Greg Stephens^c

^a The Management School, University of Bath, Bath BA2 7AY, United Kingdom

^b Lancaster Centre for Forecasting, Lancaster University Management School, Lancaster University, Lancaster LA1 4YX, United Kingdom

^c School of Information Systems, University of New South Wales, Sydney 2052, Australia

ARTICLE INFO

Article history:

Received 30 September 2009

Accepted 2 July 2010

Available online 24 July 2010

Keywords:

Restrictiveness

Guidance

Judgmental forecasting

Sales promotions

System design

ABSTRACT

Restrictiveness and guidance have been proposed as methods for improving the performance of users of support systems. In many companies computerized support systems are used in demand forecasting enabling interventions based on management judgment to be applied to statistical forecasts. However, the resulting forecasts are often 'sub-optimal' because many judgmental adjustments are made when they are not required. An experiment was used to investigate whether restrictiveness or guidance in a support system leads to more effective use of judgment. Users received statistical forecasts of the demand for products that were subject to promotions. In the restrictiveness mode small judgmental adjustments to these forecasts were prohibited (research indicates that these waste effort and may damage accuracy). In the guidance mode users were advised to make adjustments in promotion periods, but not to adjust in non-promotion periods. A control group of users were not subject to restrictions and received no guidance. The results showed that neither restrictiveness nor guidance led to improvements in accuracy. While restrictiveness reduced unnecessary adjustments, it deterred desirable adjustments and also encouraged over-large adjustments so that accuracy was damaged. Guidance encouraged more desirable system use, but was often ignored. Surprisingly, users indicated it was less acceptable than restrictiveness.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

Managers in companies often use computerized support systems to produce forecasts of demand for their products. These systems use statistical algorithms to extrapolate past patterns, but also provide facilities for managers to apply judgmental adjustments to these forecasts, where they consider these to be appropriate. However, recent research has shown that smaller adjustments tend to waste considerable management time and can also lead to reductions in forecast accuracy [1]. Often small adjustments are made because the forecaster has falsely seen patterns in the noise associated with the demand time series. In addition, such adjustments may be prompted because the forecaster has an illusion of control over the variable that is being forecast [2]. In contrast, larger adjustments *can* be effective in improving accuracy [1]. This is because they are usually made to take into account the effects of an important special event (e.g. a sales promotion campaign), which the statistical forecast has ignored. This raises the question: can facilities be incorporated into forecasting support systems to reduce the number of gratuitous (and damaging) interventions that managers make to their forecasts while, at the same time, not

discouraging (or even encouraging) them to make adjustments when these are necessary?

The information systems literature has identified two approaches that are designed to improve the way in which people use decision support systems, restrictiveness and guidance [3]. Both approaches conceptualize a gap between a descriptive model of how people use the system and a normative model [4] and they involve the incorporation of features designed to move the user towards the normative model.

System restrictiveness is defined as 'the degree to which and the manner in which a Decision Support System limits its users' decision making processes to a subset of all possible procedures' [3]. Restrictiveness can be applied to the structure of decision making and forecasting processes by restricting the set of available activities (e.g. a forecasting system could prevent the use of more complex models where their use might be damaging, simply be excluding the facility) and to the sequencing of activities (e.g. a system could require users to record a reason for adjusting a statistical forecast before being allowed to produce forecasts for other products). It can also be applied to the execution of a decision making or forecasting process. For example, it could restrict the size of smoothing parameters that can be used in exponential smoothing to a particular range to prevent forecasts overreacting to noise in a time series. Alternatively, it could proscribe behaviors that are associated with cognitive biases.

* Corresponding author. Tel.: +44 1225 383594; fax: +44 1225 386473.
E-mail address: mnsppg@bath.ac.uk (P. Goodwin).

The second approach to improving system use is decisional guidance. Silver [3] defines this as ‘the degree to which and the manner in which a Decision Support System guides its users in constructing and executing decision making processes, by assisting them in choosing and using its operators’. As Silver notes, support systems are intended to combine the strengths of human judgment with those of machines and hence a system can provide guidance on when judgmental inputs are most appropriate.

Although restrictiveness and guidance are not mutually exclusive features of support systems, the less restrictiveness there is in a system, the greater the scope, and possible need for guidance [3]. This suggests that it is useful to examine the relative effectiveness of the two approaches. While both have the objective of improving the quality of decision making or forecasting, their effectiveness also needs to be assessed on other important dimensions such as user acceptance, effect on user learning and the efficiency with which the task is carried out [5]. However, despite the attention that they have received in the literature, we know of only one empirical study that has directly compared the effectiveness of restrictiveness and guidance. In this case the two approaches were tested on undergraduate students using a knowledge-based system [6]. The authors of this paper also highlighted the absence of direct comparative studies in this area.

The current paper examines the relative effectiveness, on a number of key dimensions, of restrictiveness and decisional guidance in a support system, which was designed to improve the quality of judgmental adjustments to statistical forecasts generated by the system. This type of system is very widely used in practice [1]. In an experiment, participants in one set of treatments encountered restrictions, which prevented them from making adjustments below a certain size. Participants in a second set of treatments received guidance on whether or not they should make adjustments when the system identified that their behavior departed from that prescribed by the system. The next section reviews the relevant literature and sets out the specific research questions that were investigated. Section 3 then describes the experiment which was designed to address these questions and the results are discussed in Section 4. Finally, the paper considers the practical implications of the results and makes suggestions for further research.

2. Literature review and research questions

While some companies rely solely on managers' judgments to make forecasts (e.g. [7]), forecasting in many supply chain businesses is a semi-structured task where both statistical algorithms and human judgment can play complementary roles [8–10]. Statistical methods are efficient at detecting systematic patterns from noisy time series while humans can take into account the effects of important one-off events that are known to be occurring in the future. Performance in the forecasting task is therefore likely to be improved through the use of support systems and researchers have explored the effectiveness of a number of design features of such systems in the context of forecasting [11] and in other domains (e.g. [12,13]).

When the task involves applying managerial judgmental adjustments to statistical forecasts, two stages of the task need to be supported: (1) the decision on whether or not to adjust the statistical forecast and (2) (if a decision is made to adjust) the decision on how large the adjustment should be. Some researchers have investigated the effectiveness of facilities designed to support the second stage of the task. For example, [14] explored the use of a database of analogous special events to allow the user to estimate the effects of forthcoming events, while [15]

investigated the role of decomposition in simplifying the task of adjusting forecasts for multiple special events. However, there have been relatively few studies which have looked at ways of supporting the decision on whether or not to adjust in the first place. One exception was a study by Willemain [16], which found that adjustment is likely to improve accuracy when a naïve forecast outperforms a more sophisticated statistical forecast. Goodwin [17] found that requiring forecasters to give reasons for their adjustments significantly reduced the frequency of unnecessary adjustments without reducing the propensity to make adjustments that were beneficial. Indeed, even requiring a forecaster to answer the question: “Do you wish to make an adjustment?” reduced the propensity to make unnecessary adjustments, indicating the significant impact that small changes can make to the operation of a support system.

A major motivation for focusing on the decision to adjust is the finding that companies devote huge amounts of time and effort to the forecast adjustment process. For example [1] found that a major food processing company judgmentally adjusted 91% of its statistical forecasts, while [18] estimated that a pharmaceutical company devoted 80 person hours of management time each month to forecast review meetings where the sole objective was to agree and adjust product demand forecasts. These studies have found that much of this effort does not contribute to accuracy. For example, in the pharmaceutical company studied in [18] half the improvements obtained through adjustment improved accuracy by less than 0.37% and only 51.3% of statistical forecasts were improved through adjustment. (It should be noted that in some cases adjustments are made by managers to avoid extreme errors, rather than to improve average accuracy. For example, adjustments may be motivated by the desire to avoid individual absolute percentage errors of over a certain percentage, which may attract the attention of senior managers.)

A recent field study [1] found that the effectiveness of adjustments is partly related to their size. Larger adjustments tend to be more likely to improve accuracy because they are usually made with good reason. For example, an important event is known to be occurring in the future and the effects of this have not been incorporated into the statistical forecast. There are several possible reasons why smaller adjustments tend to be ineffective, or even damaging to forecast accuracy. Several studies [19–21] have shown that forecasters often see systematic patterns in the noise associated with time series and they make unnecessary adjustments to try to include these false patterns into the forecast. In addition, in some companies there is a tendency for forecasters to tweak forecasts merely to justify their role [1]. Smaller adjustments may also be made when forecasters are doubtful about the reliability of information about future events. As a result they may be unwilling to commit themselves to large adjustments, so instead, they ‘hedge their bet’ by compromising with a small adjustment.

These findings are consistent with the normative view of the forecasting adjustment decision, as exemplified by the Forecasting Principles project [22] that adjustments should only be made on the basis of important domain knowledge [23]. If important domain knowledge implies relatively large adjustments then this suggests that a restrictive forecasting support system that prohibits smaller adjustments will move forecasters' use of the system closer to that which is deemed to be normative. When initially deployed, such a system might not reduce wasted management effort because forecasters would still spend time estimating adjustments only to find that they are prohibited. However, as experience with using the system increases, learning should reduce these wasted interventions and eventually lead to more efficient use of management time. Nevertheless, prohibiting a user from carrying out certain tasks may have a detrimental

effect in the acceptability of the system. As Silver [3] points out, restrictiveness must not be such as to put people off using system—it should promote its use. The acceptance of systems has been the subject of extensive research, which Venkatesh and others have attempted to bring together in their Unified Theory of Acceptance and Use of Technology (UTAUT) model [24]. One of the determinants of acceptance in this model is ‘performance expectancy’, defined as ‘the degree to which an individual believes that using the system will help him or her to attain gains in job performance’. Forecasters who believe that their small adjustments are likely to improve accuracy would have low performance expectancy of such a restrictive system and, consequently, be unlikely to accept it.

Guidance is an alternative approach to supporting adjustment decisions that might be more acceptable. It can take a number of different forms [25]. Suggestive guidance recommends particular courses of actions to the user, while informative guidance provides users with relevant information, but does not suggest a course of action. Both forms of guidance can be provided either in a pre-defined form, where all users receive the same guidance under a given set of conditions, or in a dynamic form where the advice is customised to meet the apparent needs of individual users under particular conditions [5]. Furthermore, guidance, irrespective of its type, can either be delivered automatically, or only when requested. One group of researchers [25] investigated the effectiveness of providing different forms of guidance in a task that involved the selection of a forecasting model and found that, while all forms of guidance improved decision quality, suggestive guidance was most effective. It also led to the greatest user satisfaction and reduced decision time. However, it was less effective than informative guidance in fostering learning about the problem domain.

Different researchers sometimes use alternative terms when they refer to guidance. For example, it can be regarded as a form of feedback designed to enable the user to learn about how to carry out the task more effectively [26]. When invoked in response to a user's intended action, suggestive guidance explicitly contrasts this intended action with the action that the system designer has deemed to be normative.

The psychological ‘advice literature’ offers another perspective on guidance. An extensive review of this literature can be found in [27]. Although the literature is primarily oriented to advice provided by human experts, some researchers have explored the effectiveness of advice provided by machines [28]. This research has shown that while advice generally improves the quality of judgments, it is also often discounted in that people ‘[do] not follow their advisors’ recommendations nearly as much as they should have’ [27]. A number of explanations have been put forward for this. For example, it has been argued [29–31] that advice discounting partly takes place because decision makers have access to their own rationale for choosing a particular course of action while the rationale of their advisors may be less accessible. Also, imposed advice is less likely to be followed than advice which is actively solicited [27].

This discussion raises a number of questions:

- 1) Which feature of a support system, restrictiveness or guidance, is more effective in moving forecasters’ behavior closest to that of a normative approach?
- 2) Which feature leads to the most efficient decision making? Efficiency in this case is a low average time to produce forecasts.
- 3) Which feature is more effective in stimulating learning so that a movement towards normative behavior is most rapid?
- 4) Which feature is more acceptable to forecasters?

The next section describes an experiment, which was designed to address these questions.

3. Experimental design

In the instructions for the experiment participants were told that they were forecasters working for a manufacturing company, which supplies supermarkets with its products and that each month their task was to produce a forecast of total demand for a single product for the following month. Participants were told that the supermarkets sometimes run promotion campaigns and notify the manufacturer of the details of these campaigns one month ahead. Also available to the forecaster were occasional items of ‘soft’ information such as rumours or the ‘gut feel’ of some of the managers associated with the company’s operations. The nature of this soft information, and the overall design of the experiment, was based on the authors’ detailed observations of forecasting meetings and processes in several manufacturing companies [18].

The participants used an experimental computerized support system (ESS) to produce the forecasts. The ESS initially provided the following information graphically:

- i) Data on a product’s demand for the last 30 months;
- ii) statistical baseline forecasts for the last 30 months. These forecasts were derived from data that had been cleansed of estimated promotion effects and were based on simple exponential smoothing or the Holt-Winters method. They were obtained automatically by using the expert system that is incorporated in the *Forecast Pro* forecasting system [32];
- iii) a statistical baseline forecast for the next month (provided both graphically and numerically);
- iv) estimated past promotion effects;
- v) a message board giving details of any sales promotion that was due to take place next month, together with the estimated effect of recent similar promotions. To simulate a real forecasting environment; this board also occasionally displayed rumours or other managers’ speculations relating to next month’s demand. As is typical of forecasting review meetings [18] many of the manager’s views were contradictory (e.g. Sales Manager: “I’ve a gut feeling that we’ll see slightly better than expected sales next month if the mood of my sales staff is anything to go on.” Accountant: “I recall you telling me something similar a year ago and sales went down!”), and sometimes the factual information on forthcoming promotions was qualified by managerial opinion (e.g. “Price Star supermarkets are running a money-off token promotion next month. Their last 2 promotions of this type generated estimated extra sales of 45 and 57 units, respectively”). Sales Manager: “But can we trust them to display our product prominently in their stores?”).

Fig. 1 displays a typical screen from the system. After each forecast had been made the screen was updated to include the information for the next month. Forecasts were required for months 31–71. In each case, the participant had to decide whether to adjust the statistical forecast and, if a decision was made to adjust, what the new forecast should be. The new forecast could be indicated by either clicking on the graph in the appropriate place or entering the forecast into a text box.

The simulated demand time series were either ARIMA (0,1,1) or they followed a linear trend with multiplicative seasonality. In both cases the series were subject to either low noise ($\sim N(0,18.8)$) or high noise ($\sim N(0,56.4)$). The formulae used to

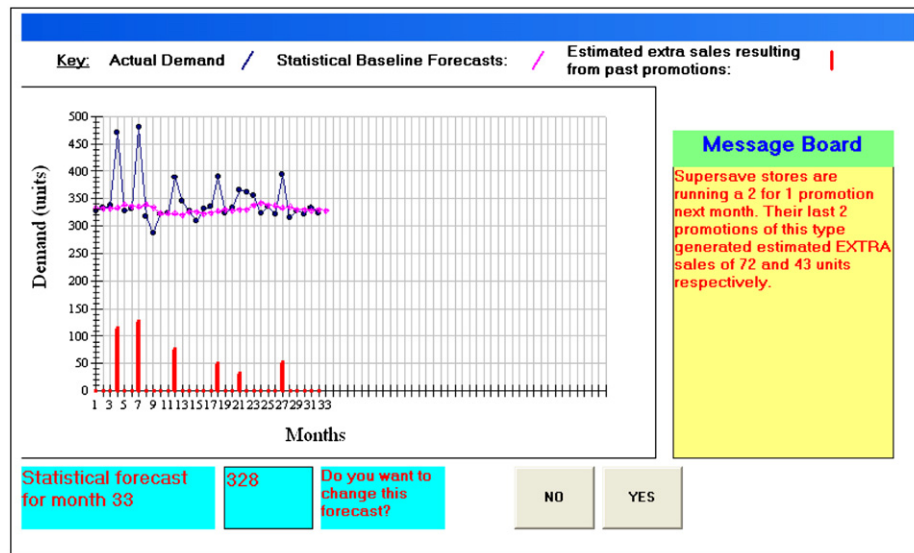


Fig. 1. A typical screen display.

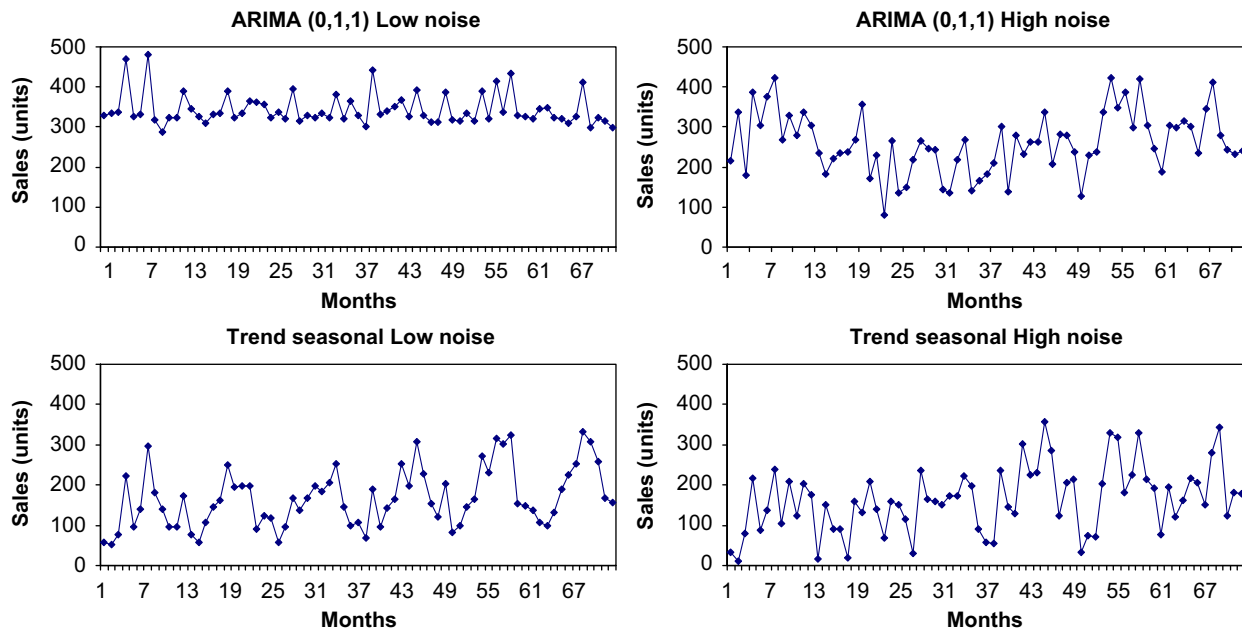


Fig. 2. Time series.

generate these series were

ARIMA(0,1,1) series : $Y_t = Y_{t-1} - 0.7e_{t-1} + e_t$

Trended seasonal series : $Y_t = (a + 1.5t)S_i + e_t$

where Y_t is the observation at time t , a the trend at $t=0$, e_t the noise at t and S_i the seasonal index for month i .

In 9 of the 41 forecast periods (promotion periods) the series was disturbed by the effects of a sales promotion. The size of the effects was roughly in line with those reported in a study, which looked into ketchup and yogurt promotions [33]. These effects had a correlation of +0.74 with the means of the estimated effects of past promotions that were displayed on the message board. The series contained no pre- or post-promotion effects. Such effects are not observable in many product sales series, particularly those where consumers are unable to stock up on the promoted product [34,35].

In the 7 periods when soft information was provided (rumour periods) this information was contradictory on three occasions. On the remaining occasions, the managers' speculations suggested a particular direction of adjustment to the statistical forecast and this conformed with the required direction of adjustment on 62.5% of occasions. Thus, adjusting on the basis of these speculations was a risky strategy. Not only was there a 37.5% probability of adjusting in the wrong direction, which would increase the forecast error, but also even if the correct direction was chosen, an adjustment that was more than twice the required adjustment would also increase the error. Fig. 2 displays the series for months 1–71.

Participants were randomly assigned to one of three treatments: control, restrictiveness or guidance. The control group performed the forecasting task without any restrictions on the size of their adjustments or any guidance on when they should make adjustments. When participants in the restrictiveness

treatment tried to make an adjustment below a pre-specified size they received a message: “The system will not allow adjustments of [x] units or less. It only permits adjustments for major factors that the statistical forecast has not allowed for.” They then had the option of trying to make a revised adjustment or accepting the statistical forecast after all. To determine x the program needed to assess whether a participant’s judgmental adjustment was likely to be an adjustment for noise (which it aims to prohibit) or one made to take into account a promotion effect (which it aims to allow). To try to discriminate between these two circumstances it compared the size of past promotion effects with an estimate of the standard deviation of the noise associated with a series. The latter was estimated using the RMSE of the first 30 baseline forecasts on the cleansed data. This yielded a promotion effect ratio, where

$$\text{Promotion effect ratio} = \frac{\text{Mean promotion effect}}{\text{Forecast RMSE}}$$

Investigation of this ratio for the low noise series suggested that a restriction prohibiting absolute adjustments of below 2 estimated noise standard deviations (31 sales units) would still allow for sufficient adjustment to take into account all the promotion effects. There is roughly only a 5% chance that a normal period observation would have required an adjustment of this size.

For the high noise series, a restriction prohibiting absolute adjustments of below one estimated noise standard deviation (i.e. 51 sales units) would still allow sufficient adjustment to take into account all the promotion effects (only one promotion effect in the past data was below the permitted adjustment level). There is roughly a 32% chance that a normal period observation would have required an adjustment of this size to yield a forecast error of zero, but obviously the discrimination between judgmental adjustments for promotion and noise effects is bound to be problematical for a high noise series. Whatever its potential limitations, this approach seemed preferable to a system, which simply prohibited absolute adjustments of below (say) 10% of the statistical forecast since such a strategy would take no account of the noise associated with a series.

If a participant in the guidance treatment indicated that they intended to adjust the statistical forecast for a non-promotion (i.e. a normal or rumour) period the following message appeared: “Are you sure that you want to change the statistical forecast? There is no promotion campaign next month and any change you make to the forecast is likely to reduce accuracy.” Conversely, if the forecaster initially chose not to adjust the statistical forecast for a promotion period the following message appeared: “You are advised to consider adjusting the statistical forecast. It cannot take into account the likely extra sales resulting from the promotion.” In both cases facilities were provided for the forecaster either to change their mind or to stay with their original decision to adjust. Note that the guidance was not solicited by participants. Also participants had to explicitly indicate that they were going against its suggestion by pressing a button labelled either: “Yes, I still want to change the forecast” or “I’ll still leave the stats forecast unchanged”.

After completing the forecasting task, participants used the computer to complete a questionnaire in which they were asked to indicate their level of agreement with the following statements using a 1 (strongly disagree) to 5 (strongly agree) scale.

- I found the forecasting system easy to use.
- I thought that the forecasting system was useful.
- I think that the forecasting system would be acceptable to forecasters in companies.
- Using this forecasting system is likely to lead to more accurate forecasts.

- I am confident that my forecasts were as accurate as they could be, given the information that was provided.

The 130 participants who took part in the experiment were graduate or final year undergraduate students in the Australian School of Business (formerly the Faculty of Commerce and Economics) of the University of New South Wales in Sydney, Australia. The final year undergraduate students were mostly from the sponsored program and had undertaken either 12 or 18 months of industrial training. Whilst the use of students in experimental research is sometimes criticized, a study by Remus [36] found that students can act as reliable proxies for managers in decision making tasks. This is particularly the case, in experiments like this, where the task mirrors the types of job undertaken in industrial placements or at an early stage in a graduate’s career. A second issue is the participants’ motivation to complete the tasks successfully. In the current experiment all participants were presented with a screen at the end of the session, which informed them of the accuracy they had achieved. The intention of this was to motivate them by providing an assessment of their judgmental skill and to foster a spirit challenge and competition between them. No monetary incentives were provided. Remus et al. [37] found that providing monetary incentives had no effect on the accuracy of forecasts produced by participants in judgmental forecasting experiments. Other studies have shown that financial rewards can lead to weaker cognitive performance (e.g. see [38]).

4. Results

4.1. Which feature of the support system, restrictiveness or guidance was most effective in moving forecasters’ behavior closest to that of a normative model?

To answer this research question data was collected on the proportion of adjustments made to the statistical forecasts in the three period types: (a) normal, (b) rumour and (c) promotion. Movement to the normative model would be signified by a smaller proportion of adjustments in non-promotion periods and the same, or a higher, proportion of adjustments in promotion periods when compared to the control group. The data was analysed using a 3 (type of support) \times 4 (series type) \times 3 (type of period) ANOVA model with type of period treated as a repeated measure. Fig. 3 shows the proportion of statistical forecasts adjusted for each mode in each type of period: (1) normal, (2) rumour and (3) promotion. The interaction depicted between support and period type is statistically significant ($F(4,236)=4.73$; $p < 0.001$). There was no significant interaction between type of support and series type.

It can be seen that, whatever the nature of the support they received, people were tempted to make an adjustment when they received a rumour, despite the high risk associated with this strategy. Nevertheless, they (correctly) made the greatest proportion of adjustments when there was a sales promotion campaign forthcoming. Both restrictiveness and guidance were successful in reducing the proportion of unnecessary adjustments in normal periods. However, restrictiveness reduced people’s propensity to make an adjustment when it was necessary in promotion periods while guidance was successful in encouraging a greater proportion of adjustments. Thus guidance appeared to be most successful in moving decision making closer to the normative model.

It is also interesting to investigate the size of adjustments made by participants. To measure these, the absolute adjustments were taken as a percentage of the statistical forecasts. The mean of these percentages yielded the mean absolute percentage

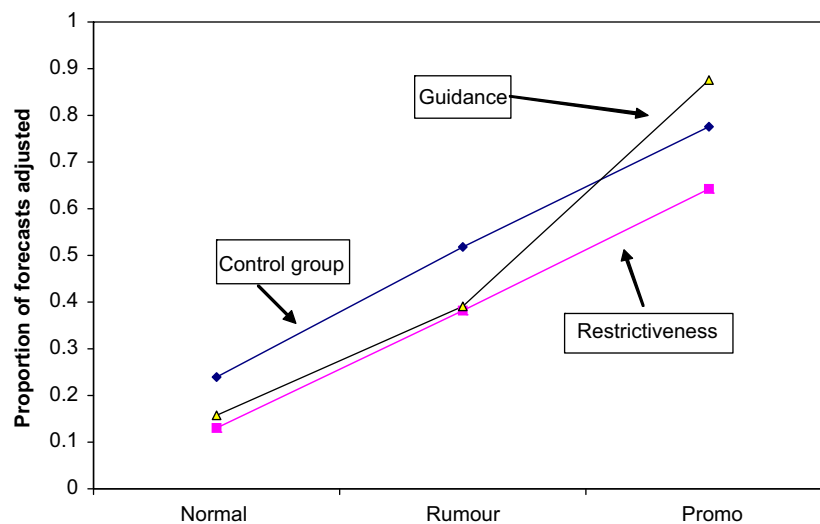


Fig. 3. Proportion of adjustments to statistical forecasts by type of period and type of support.

Table 1

Mean absolute percentage adjustment for each type of support.

Support type	MAPA
Control	23.2
Restrictiveness	31.2
Guidance	16.9

adjustment (MAPA) and Table 1 shows the MAPA for the three types of support. Note that the MAPA shown is averaged only over periods when an adjustment was made, but the relative size of the MAPAs is similar when the average is taken over all periods. There was a significant main effect for type of support ($F(2,118)=12.94$; $p < 0.0001$). Restrictiveness led to significantly larger adjustments than the other types of support. This is probably because people were trying to get round the ban on small adjustments.

What was the effect of the forecasters' behavior on accuracy? Because participants were not intended to forecast the noise in the series, accuracy was measured by comparing the forecasts with the underlying signals of the simulated time series (i.e. the demand figures before noise was added). This approach has been used in several studies including [14,17,19,39]. However, to allow comparisons with other studies, we also display the accuracy of forecasts compared to the actual demand figures in Fig. 2a. Average accuracy was measured using the mean absolute percentage error (MAPE), though similar results were obtained when the median absolute percentage error (MdAPE) was used.

When ANOVA was applied to the MAPEs there was no significant interaction between type of support and period type or type of support and series type, but there were significant main effects for both variables (for type of support: $F(2,118)=5.39$; $p < .006$ for period type: $F(2,236)=11.27$; $p < 0.0001$). Table 2a shows the MAPEs by type of support and Table 2b shows the MAPEs by period type. Most interestingly, restrictiveness actually led to a statistically significant reduction in accuracy when compared to the control. Moreover, guidance failed to yield significant improvements in accuracy over the control group. Not surprisingly, rumour periods were the least accurately forecast as they had only limited information value and were contradictory in 37.5% of cases.

A series of decision trees was constructed to obtain a more detailed assessment of how the participants reacted to the different types of support and the consequences of their decisions.

Table 2a

Mean absolute percentage errors for each type of support.

Type of support	MAPE against signal	MAPE against actual
Control	23.4	25.1
Restrictiveness	27.1	25.4
Guidance	23.4	22.4
Unadjusted statistical forecasts	22.4	22.3

Table 2b

Mean absolute percentage error against signal for each period type.

Period type	Type of support			Unadjusted statistical forecast
	Control	Restrictiveness	Guidance	
Normal	22.9	24.7	22.8	19.6
Rumour	26.3	33.5	25.4	22.4
Promotion	22.7	28.9	23.5	30.2

The trees in Fig. 4 show, for the control group, the effect of the participants' adjustment decisions on forecast accuracy. As expected, when people made adjustments in normal or rumour periods it was much more likely that these adjustments would reduce accuracy than improve it (e.g. in the control group over 70% of these forecasts were worsened through adjustment). In contrast, when people made adjustments in promotion periods, as expected, these tended to improve accuracy (over 83% of these forecasts were improved through adjustment). Thus the advice that was provided in the 'guidance' treatment, which told participants not to adjust in non-promotion periods, but to make adjustments in promotion periods was sound.

For participants in the restrictiveness treatment, data was collected on the percentage of times that they: (a) decided not to change the statistical forecast after being prohibited from making a small change or (b) simply increased the size of the adjustment after being informed that a small change was not allowed. In each case the effect of this behavior on accuracy was also measured. The decision trees in Fig. 5 show the results for the three types of periods. It can be seen that when people were banned from making a small adjustment, they often responded by making a larger adjustment, rather than simply deciding not to adjust after

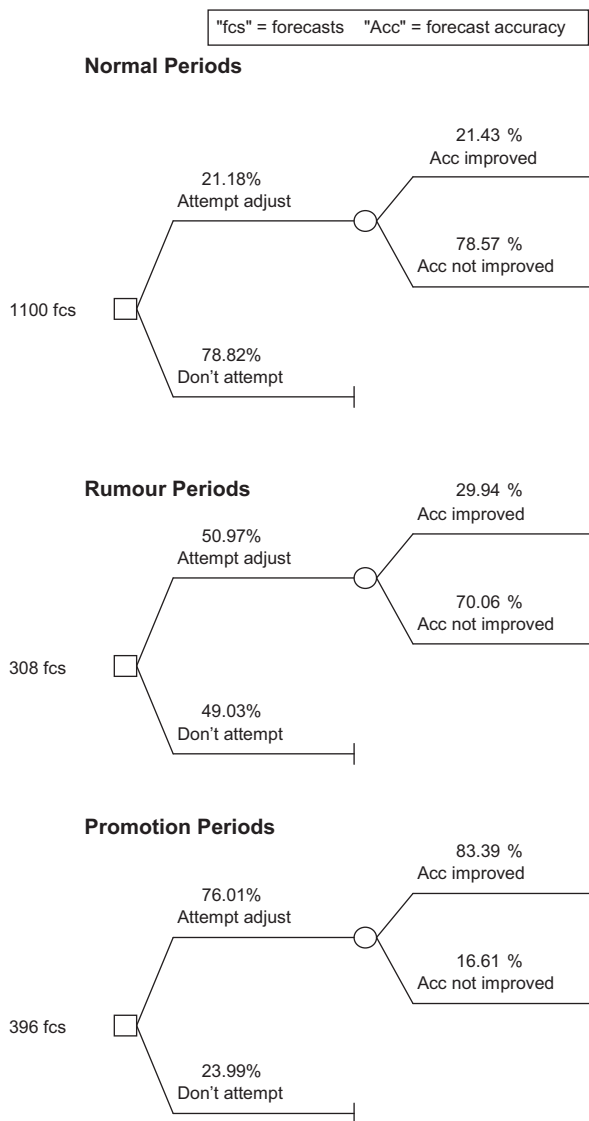


Fig. 4. Decision making in the control group.

all. This was particularly the case in 'rumour' periods where 67% of 'too small' adjustments were revised upwards and the effect was that 78.6% of these revised adjustments damaged accuracy. In normal periods people tended to be less insistent on making a larger adjustment after a small adjustment had been rejected, but when they did, 92.9% of their adjustments reduced accuracy. In promotion periods, as already reported, restrictiveness led to fewer adjustments being attempted in the first place (64.9% vs. 76.0% in the control). However, this negative effect was partly mitigated by its implicit encouragement to make larger adjustments, which led to improvements on 73.5% of occasions since larger adjustments were in fact needed here.

How did people react to the guidance they were given? As the decision trees in Fig. 6 show, in normal and rumour periods they tended to go against the good advice not to make an adjustment—they did this over 77% of the time in both cases. Recall that the advice did not simply appear in a corner of the screen where it could easily have been ignored participants attempting an adjustment had to explicitly indicate that they wanted to go against the advice before they could proceed. People were more responsive to guidance, which told them to make an adjustment (when they had initially decided not to) in promotion periods.

However, the acceptance of this advice was still surprisingly low—it was followed in only 48.2% of cases.

4.2. Which feature led to the most efficient decision making (i.e. the lowest average time to produce forecasts)?

There were no significant differences between the control group and the restrictiveness, and the guidance groups in the mean time took to produce the forecasts. Unsurprisingly, forecasts for the series with linear trend and multiplicative seasonality took significantly longer to make on average (mean time: 22 s) than those for the ARIMA series (mean time: 14 s) ($p < 0.05$).

4.3. Which feature was most effective in stimulating learning so that a movement towards normative behavior is most rapid?

Fig. 7 shows the total number of times that restrictiveness was activated as the experiment progressed for the entire group in the restrictiveness treatment. This clearly shows that people generally learned not to make small adjustments as the experiment progressed.

Fig. 7 also shows the number of occasions that guidance was activated. This curve falls less steeply than that for restrictiveness, suggesting a slower rate of learning. However, there are two differences between the features. First, restrictiveness only reminds users not to make small adjustments while guidance reminds them both not to adjust in non-promotion periods *and* to make adjustments in promotion periods so there are two things to learn. Second, the user may be prepared to activate guidance deliberately in order to prevail with an adjustment that they are determined to make. There is little point in deliberately activating restrictiveness since it will not allow the user to continue with a desired adjustment.

Of course, learning to manipulate the support system and to avoid its constraining interventions does not necessarily imply that the user is learning to carry out the task in a manner which is closer to the normative behavior intended by the system designer. To investigate this we recorded the percentage of times that participants made decisions which conflicted with the normative approach (i.e. the percentage of times that they made an adjustment in a non-promotion period or failed to adjust in a promotion period). Fig. 8 compares the results for the first 15 periods that required forecasting with the last 15 for the different types of period. The evidence that restrictiveness and guidance fostered learning towards the normative approach as the experiment progressed is weak. Even though the percentage of conflicting decisions declined as time went on for normal periods, this same phenomenon was observed in the control group. Thus the improvements might have occurred simply because the patterns in the later part of the time series offered less enticement to participants to make adjustments or because the participants were becoming fatigued. Alternatively, there may have been a natural tendency to learn, irrespective of the type of support. Thus we can only conclude that the *relative* benefits of restrictiveness and guidance were obtained early on and that there was no improvement from this initial advantage. The graph for the rumour periods also suggests no tendency to learn; it merely implies again that the advantages of restrictiveness and guidance were acquired early on. For promotion periods, the tendency of restrictiveness to discourage necessary adjustments is clearly seen. Again, the slight improvement over time is no better than that of the control group. This behavior was also reflected in the MAPEs for the different types of periods for the early and later parts of the experiment—these yielded no significant differences in any treatment. Thus, while people in

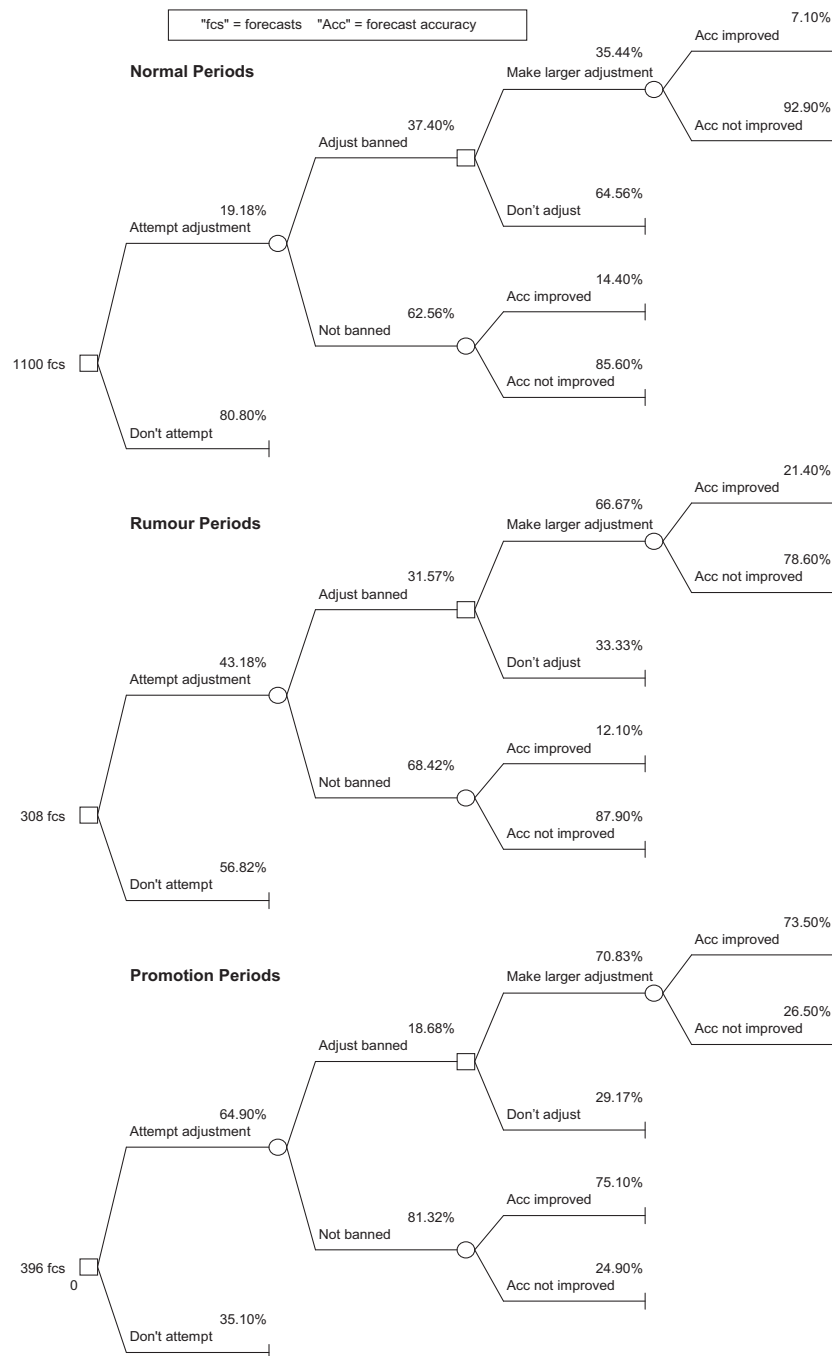


Fig. 5. The effect of restrictiveness on decision making and accuracy.

the restrictiveness and guidance groups seem to have learned over time to reduce the number of system interventions, there is little evidence that they also learned to make more accurate forecasts.

4.4. Which feature was most acceptable to forecasters?

Surprisingly, in response to the question, *I think the forecasting system would be acceptable to forecasters in companies*, participants in the restrictiveness treatment indicated a significantly higher level of agreement than those in the guidance treatment (mean=3.35 vs. 2.85, $p=0.0138$, two tail, where the scale ranges from 1=strongly disagree to 5=strongly agree). It might have

been expected that restricting participants' actions would be considered to be less acceptable. Silver [3] has suggested that people may prefer highly restrictive systems because these are easier to use and remove the burden of having to decide which features to employ. However, in this experiment, restrictiveness thwarted actions that the forecasters wished to carry out.

Apart from a weak, but significant, negative partial correlation between the total number of adjustments made by participants and their agreement with the statement: "I thought the forecasting system was useful" (after controlling for mode and series) ($r=-0.177$, $p=0.046$) none of the responses to the questionnaire were significantly correlated with the proportion of adjustments made by participants (even after controlling for mode and series). Nor were any of the dimensions measuring attitude to the system

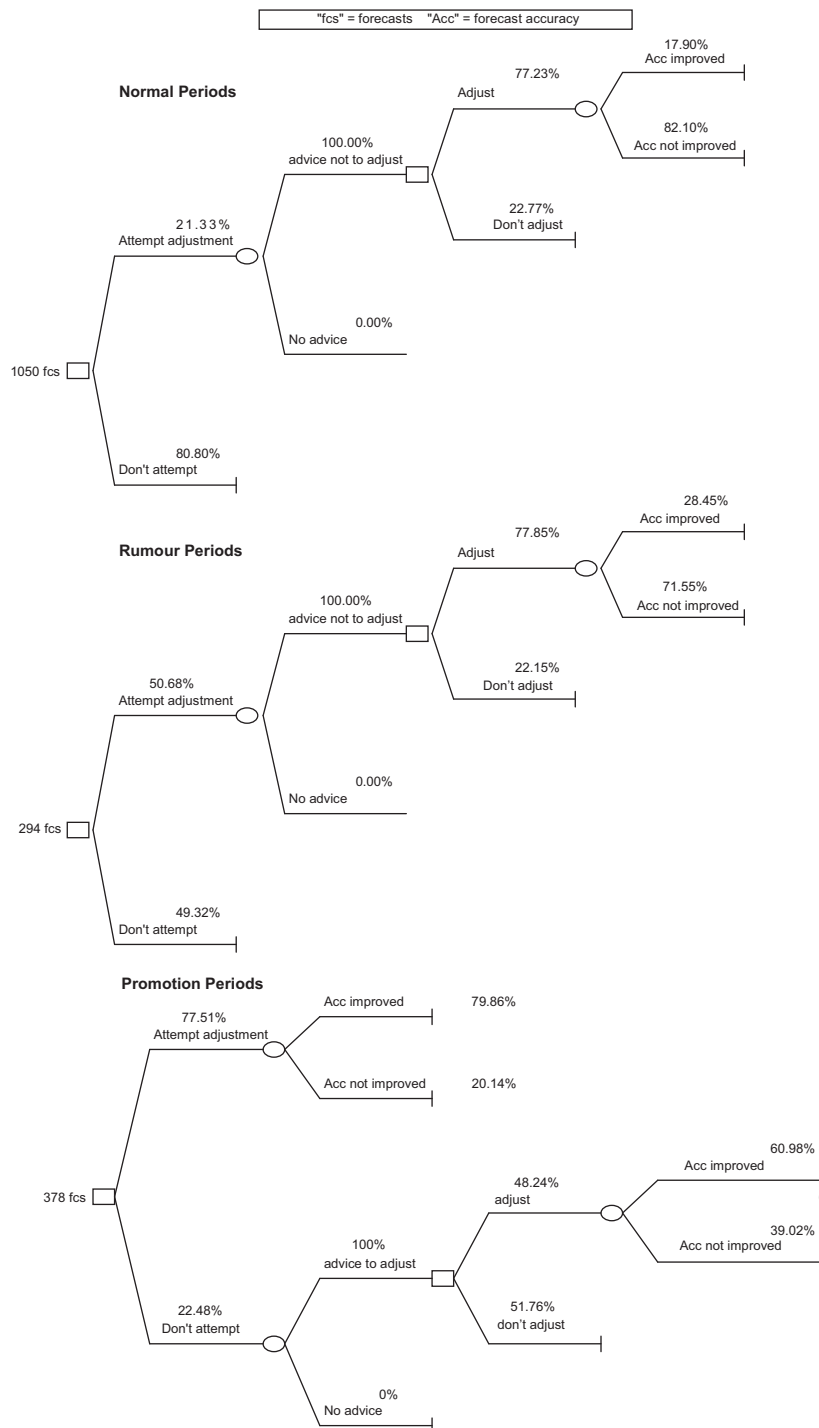


Fig. 6. The effect of guidance on decision making and accuracy.

correlated with the amount of advice received or the number of restrictions experienced.

5. Discussion

The results of this study suggest that well-intentioned design features of support systems can have damaging effects on the performance of users. People often attempted to circumvent the constraints imposed by the restrictiveness feature and the accuracy of their forecasts was reduced as a consequence. They also frequently ignored the guidance they were given—a result

which was consistent with the findings of Antony et al. [6] in their study of knowledge-based systems. Surprisingly, guidance was less acceptable than restrictiveness. It is, of course, possible that at least some of the participants had misunderstood the nature of the task or that they were treating the task either carelessly or in a deliberately negative fashion because of lack of motivation. To establish whether there was any evidence for this we first identified participants who made an excessive number of downward adjustments to the statistical forecasts in promotion periods, when, of course, an upward adjustment would normally be required. Ten participants who downwardly adjusted 4 or more of their 9 forecasts for promotion period were judged

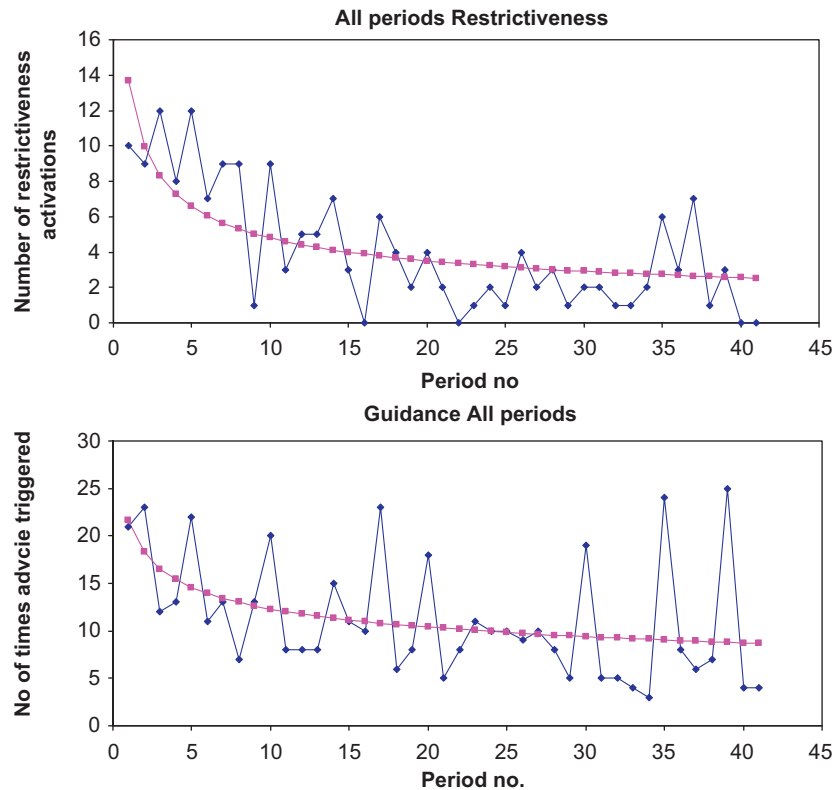


Fig. 7. Number of occasions when restrictiveness and guidance were evoked (with fitted trend lines).

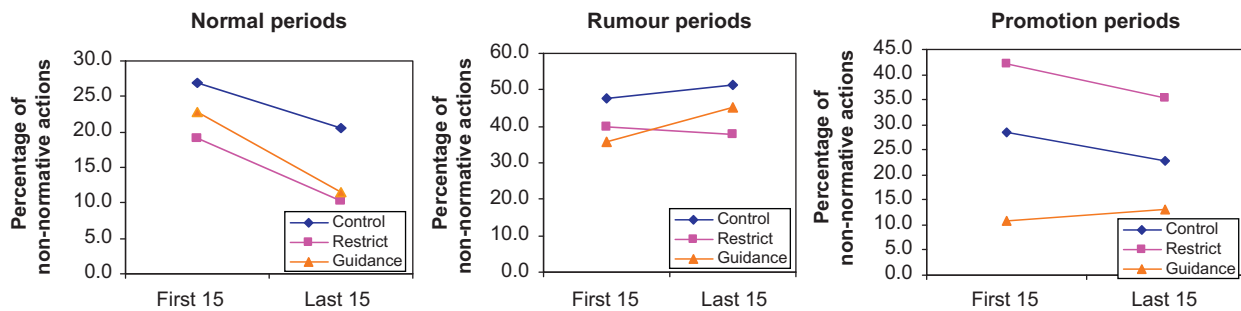


Fig. 8. Percentage of time participants' decisions conflicted with those of the normative approach in the first and last fifteen periods

to be dubious. Two additional participants never made any adjustments at all, despite the evidence that promotion periods required an upwards adjustment. However, when the analysis was repeated with these 12 participants removed (so that 118 participants remained) similar results to those reported earlier were obtained.

Although guidance was more successful in moving people's use of the system closer to that deemed to be normative, the extent to which it was ignored needs to be explained. One possibility is that guidance fared badly because it was imposed on users [27]. The relatively low level of acceptance for guidance provided some support for this. Alternatively, people may have resented having their initial decisions challenged. If this is the case providing guidance before the adjustment decision is made would be likely to be beneficial. A third possibility is that the concise form in which the guidance was conveyed did not allow it to compete with the user's internal rationale for making a given decision. The fact that advice was most often ignored in rumour periods adds weight to this latter explanation. It seems that colourful speculations or rumours that have low reliability

(e.g. Period 47: "Haven't they predicted that we'll have heavy rain next month? These medium term weather forecasts might not be that reliable, but rain will hit our sales.") are likely to have far greater salience than a piece of terse advice cautioning against adjustment in a non-promotion period. This mirrors Tversky and Kahneman's [40] finding that judges will ignore statistical base-rates in favour of anecdotal case-based information that has little reliability. Of course in real organizational environments forecasters may also feel obliged to act on rumours, especially if they emanate from more senior managers. The beneficial effects of guidance, such as they were, were also achieved early on and there was only weak evidence of an improvement from this position as time went on. Parikh et al. [5] found that suggestive guidance was less effective than informative guidance in developing user learning about the problem domain. To foster learning it might therefore be worth coupling the suggestive guidance with some form of feedback, which either provided direct information about the task to support the advice and/or provided information showing the past consequences of ignoring or adhering to the advice.

The main intention of restrictiveness was to dissuade participants from adjusting forecasts in non-promotion periods. While it achieved this objective to an extent, the resulting benefits were outweighed by two associated disadvantages. First, once people decided to make an adjustment in non-promotion periods, they often seemed determined to persevere and simply increased the size of the adjustment to get round the restriction, thereby tending to exacerbate the inaccuracy of the forecast. Second, ironically, restrictiveness appeared to reduce people's propensity to make the adjustments that were required in promotion periods. The result was that, overall, restrictiveness significantly reduced the accuracy of forecasts when compared with the control group. Again it would be interesting to see if these negative effects could be mitigated if restrictiveness was coupled with some form of feedback. It was also expected that restrictiveness would eventually improve the efficiency of decisions by discouraging participants from wasting time pondering over small adjustments that would serve no purpose. This expectation was also not borne out by the results. Although those in the restrictiveness group attempted fewer adjustments, this did not significantly reduce the mean time they spent on the forecasting task.

6. Conclusions

The main conclusion of this paper is that neither restrictiveness nor guidance were wholly successful in fostering improved use of a support system. Indeed, restrictiveness, though more acceptable to forecasters than guidance, actually encouraged counter-productive behavior and significantly reduced forecast accuracy. Although guidance was more effective in encouraging decisions that were closer to the normative approach its impact was limited because it was frequently ignored. The experiment illustrates the danger that well-intentioned design features in support systems can be resisted by users with potentially detrimental effects on the quality of their decision making.

Despite this the experiment provided enough evidence to suggest that support systems are worth pursuing and that the ideas of guidance and restrictiveness could be worthwhile features of such systems if their disadvantages can be avoided. They both have the potential to deliver improved forecasts. Providing feedback in association with these facilities might be one way of achieving this and would be a worthwhile avenue for future research. Future experiments could also usefully investigate the role of restrictiveness and guidance when forecasts of demand are provided for multiple time horizons and where minimum allowable adjustments are specified in advance, with adjustments possibly being expressed as percentages rather than as absolute values. The effect of requiring managers to make adjustments *prior* to seeing the statistical forecasts (as recommended by [22]) with further adjustments being prohibited thereafter is also worth investigating.

Acknowledgements

This research was supported by Engineering and Physical Sciences Research Council (EPSRC) Grants GR/60198/01 and GR/60181/01.

References

[1] Fildes R, Goodwin P, Lawrence M, Nikolopoulos K. Effective forecasting and judgmental adjustments: an empirical evaluation and strategies for improvement in supply-chain planning. *International Journal of Forecasting* 2009;5: 3–23.

[2] Kottemann JE, Davis FD, Remus WE. Computer-assisted decision making: performance, beliefs, and the illusion of control. *Organizational Behavior and Human Decision Processes* 1994;57:26–37.

[3] Silver M. Decision support systems: directed and nondirected change. *Information Systems Research* 1990;1:47–70.

[4] Stabell CB. A decision-orientated approach to building DSS. In: Bennett JL, editor. *Building decision support systems*. Reading, MA: Addison-Wesley; 1983.

[5] Parikh M, Fazlollahi MB, Verma S. The effectiveness of decisional guidance: an empirical investigation. *Decision Sciences* 2001;32:303–31.

[6] Antony S, Batra D, Santhanam R. The use of a knowledge-based system in conceptual data modelling. *Decision Support Systems* 2005;41:176–88.

[7] Sanders NR, Graman GA. Quantifying costs of forecast errors: a case study of the warehouse environment. *Omega* 2009;37:116–25.

[8] Blattberg RC, Hoch SJ. Database models and managerial intuition: 50% Model + 50% manager. *Management Science* 1990;36:887–99.

[9] Goodwin P. Integrating management judgment and statistical methods to improve short term forecasts. *Omega* 2002;30:127–35.

[10] Keen PGW, Morton MSScott. *Decision support systems: an organizational perspective*. Reading, UK: Addison-Wesley; 1978.

[11] Fildes R, Lawrence M, Goodwin P. The design features of forecasting support systems and their effectiveness. *Decision Support Systems* 2006;42: 351–361.

[12] Fagerholt K, Christiansen M, Hvattum LM, Johnsen TAV, Vabø TJ. A decision support methodology for strategic planning in maritime transportation. *Omega* 2010;38:465–74.

[13] Tütüncü GY, Carreto CAC, Baker BM. A visual interactive approach to classical and mixed vehicle routing problems with backhauls. *Omega* 2009;37: 138–54.

[14] Lee WY, Goodwin P, Fildes R, Nikolopoulos K, Lawrence M. Providing support for the use of analogies in demand forecasting tasks. *International Journal of Forecasting* 2007;23:377–90.

[15] Webby W, O'Connor M, Edmundson B. Forecasting support systems for the incorporation of event information: an empirical investigation. *International Journal of Forecasting* 2005;21:411–23.

[16] Willemain TR. The effect of graphical adjustment on forecast accuracy. *International Journal of Forecasting* 1991;7:151–4.

[17] Goodwin P. Improving the voluntary integration of statistical forecasts and judgment. *International Journal of Forecasting* 2000;16:85–9.

[18] Goodwin P, Fildes R, Lee WY, Nikolopoulos K, Lawrence M. Understanding the use of forecasting systems: an interpretive study in a supply-chain company. University of Bath Management School Working Paper, 2007.

[19] Goodwin P, Fildes R. Judgmental forecasts of time series affected by special events: does providing a statistical forecast improve accuracy? *Journal of Behavioral Decision Making* 1999;12:37–53.

[20] Harvey N. Why are judgments less consistent in less predictable task situations? *Organizational Behavior and Human Decision Processes* 1995;63:247–63.

[21] Armstrong JS. *Principles of Forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers; 2001.

[22] O'Connor M, Remus W, Griggs K. Judgmental forecasting in times of change. *International Journal of Forecasting* 1993;9:163–72.

[23] Sanders N, Ritzman L. Judgmental adjustments of statistical forecasts. In: Armstrong JS, editor. *Principles of forecasting: a handbook for researchers and practitioners*. Norwell, MA: Kluwer Academic Publishers; 2001. [Chapter 13].

[24] Venkatesh V, Morris MG, Davis GB, Davis FD. User acceptance of information technology: towards a unified view. *MIS Quarterly* 2003;27:425–78.

[25] Montazemi AR, Wang F, Nainar SMK, Bart CK. On the effectiveness of decisional guidance. *Decision Support Systems* 1996;18:181–98.

[26] O'Connor M, Remus W, Lim K. Improving judgmental forecasts with judgmental bootstrapping and task feedback support. *Journal of Behavioral Decision Making* 2005;18:247–60.

[27] Bonaccio S, Dalal RS. Advice taking and decision making: an integrative literature review, and implications for organizational sciences. *Organizational Behavior and Human Decision Processes* 2006;101:127–51.

[28] Wærn Y, Ramberg R. People's perception of human and computer advice. *Computers in Human Behavior* 1996;12:17–27.

[29] Yaniv I. The benefit of additional opinions. *Current Directions in Psychological Science* 2004;13:75–8.

[30] Yaniv I. Receiving other people's advice: influence and benefit. *Organizational Behavior and Human Decision Processes* 2004;93:1–13.

[31] Yaniv I, Kleinberger E. Advice taking in decision making: egocentric discounting and reputation formation. *Organizational Behavior and Human Decision Processes* 2000;83:260–81.

[32] Stellwagen EA, Goodrich RL. *Forecast pro for Windows*. Belmont, MA: Business Forecast Systems Inc.; 1994.

[33] Ailawadi KL, Neslin SA. The effect of promotion on consumption: buying more and consuming it faster. *Journal of Marketing Research* 1998;35:390–8.

[34] Hendel I, Nevo A. The post promotion dip puzzle. What do the data have to say? *Quantitative Marketing and Economics* 2003;1:409–24.

[35] Neslin SA, Stone LGS. Consumer inventory sensitivity and the post-promotion dip. *Marketing Letters* 1996;7:77–94.

[36] Remus W. Graduate students as surrogates for managers in experiments on business decision making. *Journal of Business Research* 1986;14: 19–25.

- [37] Remus W, O'Connor M, Griggs K. The impact of incentives on the accuracy of subjects in judgmental forecasting experiments. *International Journal of Forecasting* 1998;14:515–22.
- [38] Eroglu C, Croxton KL. Biases in judgmental adjustments of statistical forecasts: the role of individual differences. *International Journal of Forecasting* 2010;26:116–33.
- [39] Harvey N, Bolger F. Graphs versus tables: effects of data presentation format on judgemental forecasting. *International Journal of Forecasting* 1996;12: 119–137.
- [40] Tversky A, Kahneman D. Judgment under uncertainty: heuristics and biases. *Science* 1974;185:1124–31.